

Quantum Simpson's Paradox and High Order Bell-Tsirelson Inequalities

Yaoyun Shi*

Department of Electrical Engineering and Computer Science,
University of Michigan, 2260 Hayward Street, Ann Arbor, MI 48109-2121, USA
(Dated: December 28, 2011)

The well-known Simpson's Paradox, or Yule-Simpson Effect, in statistics is often illustrated by the following thought experiment: A drug may be found in a trial to increase the survival rate for both men and women, but decrease the rate for all the subjects as a whole. This paradoxical reversal effect has been found in numerous datasets across many disciplines, and is now included in most introductory statistics textbooks. In the language of the drug trial, the effect is impossible, however, if *both* treatment groups' survival rates are higher than *both* control groups'. Here we show that for quantum probabilities, such a reversal remains possible. In particular, a "quantum drug", so to speak, could be *life-saving* for both men and women yet *deadly* for the whole population. We further identify a simple inequality on conditional probabilities that must hold classically but is violated by our quantum scenarios, and completely characterize the maximum quantum violation. As polynomial inequalities on entries of the density operator, our inequalities are of degree 6.

PACS numbers: 03.65.Ud, 03.65.Ta, 03.67.-a, 02.50.-r

Simpson [1] used a thought clinical trial of a drug to illustrate the possibility that an association found in subgroups of a population may disappear, or even reversed, on the population as a whole. Denote by R_t^f , R_t^m , R_t , R_c^f , R_c^m , and R_c the recovery rates of the female, the male, and the combined treatment groups, respectively. As depicted in Box C1 of Fig. 1, $R_t^f > R_c^f$, $R_t^m > R_c^m$, indicating a beneficial effect on both the females and the males. However, for the combined control and treatment groups, $R_t = R_c$, erasing the beneficial effect. With small modifications as shown in C2, $R_t < R_c$, reversing the beneficial effect. Numerous real-life datasets from many fields such as epidemiology, social sciences, psychology, and sports, have been found to exhibit this reversal. The result, called Simpson's Paradox or the Yule-Simpson Effect, among several other names, is included in most introductory statistics textbooks.

To allow a generalization to quantum probabilities, we discuss Simpson's Paradox in the following model. Define a (*measurement*) *scenario* $\mathcal{M} \stackrel{\text{def}}{=} (|\phi\rangle, G, E, R)$, where $|\phi\rangle$ is a quantum state, and G , E , R are two-outcome measurements on the state space of $|\phi\rangle$. For convenience of exposition, we refer to G , E , and R as the Gender, the Treatment, and the Result measurements, respectively, and name the outcomes as **Female/Male** (F/M), **Untreated/Treated** (U/T), and **Alive/Dead** (A/D), correspondingly.

Given \mathcal{M} , consider two experiments. The first is to measure Gender, followed by Treatment, and finally Result. The second is the same, except *not* measuring Gender. The two experiments define the conditional probabilities R_t , R_t^f , etc. discussed above. For example, $R_t \stackrel{\text{def}}{=} \Pr(A|T)$, $R_t^f \stackrel{\text{def}}{=} \Pr(A|FT)$, etc.

A classical scenario is one where the measurements *commute*, i.e., changing the order of applying the measurements does not alter the outcome statistics. Conse-

quently, R_t can lie anywhere between R_t^f and R_t^m (and likewise for R_c). More precisely,

$$R_t = \alpha_f R_t^f + \alpha_m R_t^m, \quad (1)$$

where α_f (α_m) is the fraction of female (male, respectively) subjects in the combined treatment group, i.e., $\Pr(F|T)$ (or $\Pr(M|T)$, respectively). We refer to this relation as the *convexity property*. When the "treatment interval" delimited by R_t^f and R_t^m *intersects* with the "control interval" delimited by R_c^f and R_c^m , a reversal becomes possible (and requires that the gender distributions are different). Conversely, when the two intervals are *disjoint*, i.e., $R_t^f, R_t^m > R_c^f, R_c^m$, then no reversal is possible. Such is the case depicted in CI, as well as in C3 when $R_t^f = R_t^m = 99\%$, and $R_c^f = R_c^m = 33\%$.

In sharp contrast, we show that for non-classical scenarios, a reversal under disjoint rate intervals remains possible. Q1 in Fig. 1 shows that, with $R_t^f = R_t^m = 99\%$ and $R_c^f = R_c^m = 33\%$, R_t can be reduced to 0 and R_c increased to 50%, as opposed to in the classical case remaining at 99% and 33%, respectively. In the more extreme quantum scenario Q2, "taking the drug", so to speak, increases the survival rate of both gender groups from almost 0 to almost 1, yet reduces the rate for the combined groups to 0.

Such a reversal is inherently quantum as the convexity property must fail. Denote by $d_t = R_t^f + R_t^m - R_t$, $d_c = R_c^f + R_c^m - R_c$ and $S = d_t - d_c$. For classical scenarios, the convexity property implies that $|S| \leq 1$. Therefore, the value of $|S|$ exceeding 1 in quantum scenarios quantifies intuitively the extent of an inherently quantum reversal. Q1 has $S \approx 7/6$ while Q2 has S arbitrarily close to 2. We shall prove that 2 is precisely the quantum limit, i.e., $|S| < 2$ for all quantum scenarios. In particular, a maximum violation of the convexity property cannot occur simultaneously for R_t and R_c .

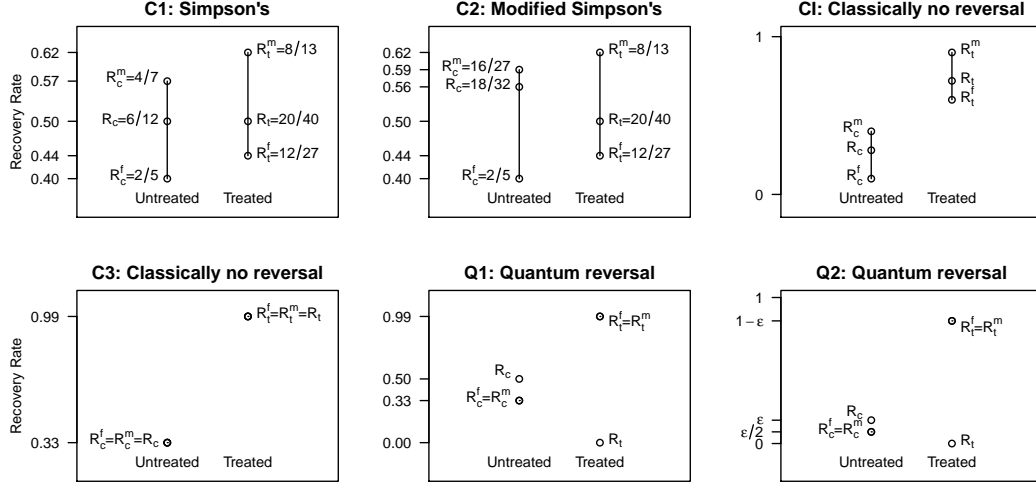


FIG. 1: **The survival rates in scenarios for the classical and quantum Simpson's Paradox.** Simpson's original example C1 shows that the positive treatment effect among the males and the females disappears when the two groups are combined. Each fraction is the number of surviving subjects divided by the total number of subjects in the corresponding group. Changing the subject numbers in the male treatment group leads to a reversal C2. Q1 and Q2 demonstrate classically impossible quantum reversals, as in both scenarios, the treatment interval is far higher the control interval, yet the combined treatment group's rate falls below that for the combined control group. The line segments depict rate intervals. The convexity property fails simultaneously for the treatment and the control groups. The classical scenario C3 is to contrast with Q1 as they behave the same on all subgroups but C3's combined rates do not change due to the convexity property. Q2 holds for all sufficiently small $\epsilon > 0$ with $O(\epsilon^2)$ precision. It represents a family of maximum quantum reversal, as $S \approx 2 - 2\epsilon \rightarrow 2$.

Before we present the details for the construction and the proof, we sketch below the underlying intuitions. A fundamental feature of quantum mechanics is that measuring a quantum system may in general *change* the state of the system. Furthermore, two different quantum measurements may be *non-commuting*, a consequence of which is that the state change incurred by one measurement would alter the outcome statistics of subsequently applying the other. This is the critical property that underlies our constructions, as well as the Heisenberg Uncertainty Principle, and other well-known quantum paradoxes such as the EPR Paradox [2] and the GHZ Paradox [3].

In our construction, the Gender and the Treatment measurements are non-commuting (though both commute with the Result measurement). Thus adding the Gender measurement may change dramatically the outcome statistics of the subsequent measurements. More specifically, the extremal violation of the convexity property $d_t \rightarrow 2$ is a consequence of the following features (illustrated in Fig. 2): (1) The **Treatment** portion of the quantum state has a small amplitude in the **Dead** subspace, but a *vanishing* amplitude in the **Alive** subspace. This implies $R_t = 0$. (2) After measuring Gender, both eigenstates retain a tiny amplitude in the **Dead** subspace, but a much larger, even though still small, amplitude in the **Alive** subspace. This gives $R_t^f = R_t^m \rightarrow 1$, and with (1) $d_t \rightarrow 2$. Q1 and Q2 are constructed so that $d_t \rightarrow 2$

and at the same time d_c is made small. We note that unlike in the violation of spatial Bell inequalities, entanglement is perhaps not relevant in our setting.

That $|S| < 2$ follows from the necessary tradeoff between d_t and d_c . The tradeoff is most intuitive when considering why extremal violation of the convexity property cannot take place simultaneously on d_t and d_c (in opposite directions). An extremal violation on d_t require the above two features (though “vanishing” can be replaced by “very small”). Consequently, $\Pr(AT)$ and $\Pr(DT)$ are small. Similarly, that $d_c \rightarrow -2$ implies $\Pr(AU)$ and $\Pr(DU)$ are also small, contradicting that the sum of those four probabilities is 1.

We now discuss the implications of our results and compare them with related works. Simpson's Paradox is important for revealing the pitfalls of drawing causal conclusions from partitioned data, and for guiding statistics-based decision making [4]. Our quantum reversal effect can play a similar role when examining quantum measurement data. With the rapid advances in experimental quantum science and engineering, it is anticipated that quantum technologies will be widely used for research and applications in the future. Thus there will be more contexts of scientific and practical importance to which our result is directly relevant.

Our inequalities share the same features of the celebrated Bell Inequalities [5, 6] and Tsirelson Inequalities [7] for differentiating classical and quantum proba-

bilities, and for bounding the latter, respectively. However, a quantum system is measured only once in Bell-Tsirelson inequalities, while ours involve repeated measurements. Bell-Tsirelson type inequalities on outcomes of repeated measurements have been discussed by many authors in the framework introduced by Leggett and Garg [8], and the related framework of “quantum entanglement in time” introduced by Brukner et al. [9]. Those inequalities are referred to as *temporal*, in contrast to the traditional, or *spatial*, inequalities. They bound quantities that are linear in the outcome correlations of repeated measurements. Consequently, they are on *linear* functions of entries of the density operator, just as the spatial Bell-Tsirelson inequalities. In contrast, our inequalities, in an equivalent form, bound polynomials of higher order (degree 6).

Thus, if one interprets broadly the terms Bell and Tsirelson Inequalities as inequalities bounding functions of the measurement outcome probabilities for classical, and quantum, respectively, models, our inequalities may be appropriately called high order Bell-Tsirelson inequalities. We note that quadratic Bell-Tsirelson inequalities were proved by Uffink [10] in a different context. We emphasize that the value of such high order inequalities does not lie in the degree per se, but in that they may, as in our case, provide a different, informative, and intuitive understanding of quantum effect.

We speculate that a quantum reversal may also occur in some *classical* settings such as human decision making. For example, in the well-known “Disjunction Effect” [11] experiment, the human subjects overall preferred Decision A than B on learning *either* value of a two-outcome random variable W . Paradoxically, B was preferred instead statistically when W ’s value was *not* revealed even though the subject knew it was already determined. The effect can be interpreted as a consequence of a violation of the convexity property, similar to our quantum reversal. Note that a “quantum” reversal in such a setting does not contradict our earlier conclusion of classical impossibility as the subject’s preference does not have a definite value. This indefinite nature resembles that of the outcome of quantum measurements, or in the language used by EPR [2], the non-existence of “elements of reality”. Indeed, the nascent area of *quantum cognition* [12] was partly motivated by the Disjunction Effect to use quantum probabilities for modeling human cognitive processes.

Construction. For two quantum states $|\alpha\rangle$ and $|\beta\rangle$, denote by $|(\alpha \pm \beta)\rangle$ the state $\frac{1}{\sqrt{2}}(|\alpha\rangle \pm |\beta\rangle)$. We will use nested parentheses but may omit the outer-most pair. For example $|\alpha + (\beta - \gamma)\rangle \equiv \frac{1}{\sqrt{2}}(|\alpha\rangle + \frac{1}{\sqrt{2}}(|\beta\rangle - |\gamma\rangle))$.

Set $H = V \otimes W$, where V has dimension 4 with an orthonormal basis $\{|t\rangle, |u_0\rangle, |u_1\rangle, |u_2\rangle\}$, and W has dimension 2 with an orthonormal basis $\{|a\rangle, |d\rangle\}$. The measurement R acts only on W , while G and E act on V only.

Denote by R^A the **Alive** eigenspace, or the projection to this eigenspace, of R , and similarly define R^F , E^T , etc. The measurements are defined through their eigenspaces

$$R^A \stackrel{\text{def}}{=} \text{span}\{|a\rangle\}, \quad R^D \stackrel{\text{def}}{=} \text{span}\{|d\rangle\}, \\ E^T \stackrel{\text{def}}{=} \text{span}\{|t\rangle\}, \quad E^U \stackrel{\text{def}}{=} \text{span}\{|u_i\rangle : 0 \leq i \leq 2\}.$$

To define G , we first define the following states:

$$|f_0\rangle \stackrel{\text{def}}{=} |(u_0 + u_1) + t\rangle, \quad |f_1\rangle \stackrel{\text{def}}{=} |u_2\rangle, \quad (2)$$

$$|m_0\rangle \stackrel{\text{def}}{=} |(u_0 + u_1) - t\rangle, \quad |m_1\rangle \stackrel{\text{def}}{=} |u_0 - u_1\rangle. \quad (3)$$

Now define

$$G^F \stackrel{\text{def}}{=} \text{span}\{|f_0\rangle, |f_1\rangle\} \text{ and } G^M \stackrel{\text{def}}{=} \text{span}\{|m_0\rangle, |m_1\rangle\}.$$

For $p, q \geq 0$, define

$$|\phi_a\rangle \stackrel{\text{def}}{=} \sqrt{p}|u_0 + u_1\rangle, \text{ and} \quad (4)$$

$$|\phi_d\rangle \stackrel{\text{def}}{=} |(u_0 - u_1) + u_2\rangle + \sqrt{q}|t\rangle. \quad (5)$$

Finally, define (the unnormalized)

$$|\phi\rangle \stackrel{\text{def}}{=} |\phi_a\rangle \otimes |a\rangle + |\phi_d\rangle \otimes |d\rangle. \quad (6)$$

By direct computation,

$$P(D|T) = 1, \quad (7)$$

$$P(A|U) = p/(1+p), \quad (8)$$

$$P(A|TF) = P(A|TM) = p/(p+q), \quad (9)$$

$$P(D|UF) = P(D|UM) = (2+q)/(2+p+q). \quad (10)$$

When $p = 1$, $q = \epsilon \rightarrow 0$, we have Q1. When $q = p^2$ and $p = \epsilon \rightarrow 0$, we have Q2. Those two examples are illustrated in Fig. 2.

Bounding the quantum violation. We prove here $|S(\mathcal{M})| < 2$ for all measurement scenario \mathcal{M} . It suffices to prove that $S' \stackrel{\text{def}}{=} S + 3 < 5$. Without loss of generality we assume that the measurements are all projective (as otherwise we can replace each POVM by a projective measurement on the extended Hilbert space). To simplify notation, we define $\ell_{DT} \stackrel{\text{def}}{=} \|R^D E^T |\phi\rangle\|$, $\ell_{ATF} \stackrel{\text{def}}{=} \|R^A E^T G^F |\phi\rangle\|$, etc. and,

$$\alpha \stackrel{\text{def}}{=} \frac{\ell_{AT}}{\ell_{DT}}, \quad \alpha_F \stackrel{\text{def}}{=} \frac{\ell_{DTF}}{\ell_{ATF}}, \quad \alpha_M \stackrel{\text{def}}{=} \frac{\ell_{DTM}}{\ell_{ATM}}, \quad (11)$$

$$\beta \stackrel{\text{def}}{=} \frac{\ell_{DU}}{\ell_{AU}}, \quad \beta_F \stackrel{\text{def}}{=} \frac{\ell_{AUF}}{\ell_{DUF}}, \quad \beta_M \stackrel{\text{def}}{=} \frac{\ell_{AUM}}{\ell_{DUM}}. \quad (12)$$

Then

$$S' = (1 + \alpha^2)^{-1} + (1 + \alpha_F^2)^{-1} + (1 + \alpha_M^2)^{-1} + \\ (1 + \beta^2)^{-1} + (1 + \beta_F^2)^{-1} + (1 + \beta_M^2)^{-1}. \quad (13)$$

By the triangle inequality,

$$\ell_{DT} \leq \ell_{DTF} + \ell_{DTM} = \alpha_F \ell_{ATF} + \alpha_M \ell_{ATM} \quad (14)$$

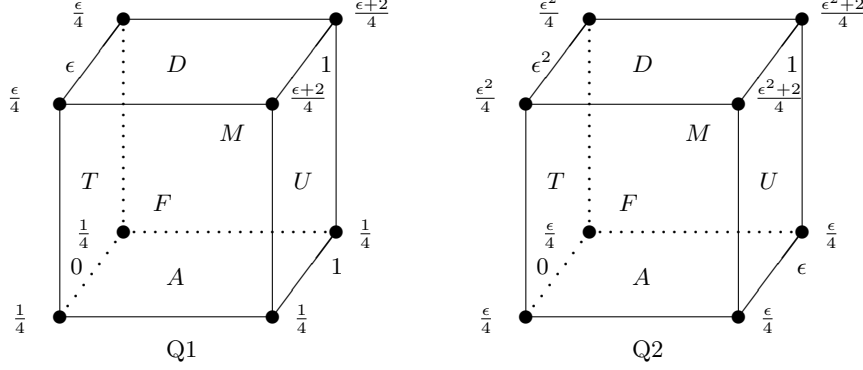


FIG. 2: Each of the three axes corresponds to one of the three measurements and each face is annotated with a measurement outcome. Each face, edge, and vertex represent the resulting state after one, two, or three, respectively, measurements, for the measurement outcomes determined by the faces incident to it, and annotated with the squared lengths (for those projections in S). For example, Q1's lower-left-vertex represents $R^A E^T G^F |\phi\rangle$ and has a squared length $1/4$.

and,

$$\ell_{AU} \leq \ell_{AUF} + \ell_{AUM} = \beta_F \ell_{DUF} + \beta_M \ell_{DUM}. \quad (15)$$

Decomposing $1 = \|\phi\|^2$ in two ways, we have

$$(1 + \alpha^2) \ell_{DT}^2 + (1 + \beta^2) \ell_{AU}^2 \quad (16)$$

$$= (1 + \alpha_F^2) \ell_{ATF}^2 + (1 + \alpha_M^2) \ell_{ATM}^2 + (1 + \beta_F^2) \ell_{DUF}^2 + (1 + \beta_M^2) \ell_{DUM}^2. \quad (17)$$

If $\ell_{DT} = 0$ or $\ell_{AU} = 0$, $S' \leq 5$. It is straightforward to verify that $S' \neq 5$ in either case. Thus $S' < 5$. Consider now that $\ell_{DT}, \ell_{AU} \neq 0$. Suppose that

$$1 + \alpha^2 \geq (1 + \alpha_F^2) \frac{\ell_{ATF}^2}{\ell_{DT}^2} + (1 + \alpha_M^2) \frac{\ell_{ATM}^2}{\ell_{DT}^2}. \quad (18)$$

Then

$$(1 + \alpha^2)^{-1} + (1 + \alpha_F^2)^{-1} + (1 + \alpha_M^2)^{-1} \leq \left((1 + \alpha_F^2) \frac{\ell_{ATF}^2}{\ell_{DT}^2} + (1 + \alpha_M^2) \frac{\ell_{ATM}^2}{\ell_{DT}^2} \right)^{-1} + (1 + \alpha_F^2)^{-1} + (1 + \alpha_M^2)^{-1} \quad (19)$$

$$\leq 2. \quad (20)$$

The second inequality follows by optimizing over all $\frac{\ell_{ATF}}{\ell_{DT}}$ and $\frac{\ell_{ATM}}{\ell_{DT}}$ that satisfy (14). Thus $S' \leq 5$. A direct computation shows that equality cannot hold. Thus $S' < 5$.

We need only consider the case when (18) fails. Rearranging Eqn. (16),

$$\begin{aligned} & \ell_{DT}^2 \left((1 + \alpha^2) - (1 + \alpha_F^2) \frac{\ell_{ATF}^2}{\ell_{DT}^2} - (1 + \alpha_M^2) \frac{\ell_{ATM}^2}{\ell_{DT}^2} \right) \\ &= \ell_{AU}^2 \left(-(1 + \beta^2) + (1 + \beta_F^2) \frac{\ell_{DUF}^2}{\ell_{AU}^2} + (1 + \beta_M^2) \frac{\ell_{DUM}^2}{\ell_{AU}^2} \right). \end{aligned} \quad (21)$$

Thus

$$1 + \beta^2 > (1 + \beta_F^2) \frac{\ell_{DUF}^2}{\ell_{AU}^2} + (1 + \beta_M^2) \frac{\ell_{DUM}^2}{\ell_{AU}^2}. \quad (22)$$

The proof that $S' < 5$ is similar as for the case when (18) holds. This completes the proof that $S' < 5$, thus $|S(\mathcal{M})| < 2$, for all measurement scenario \mathcal{M} .

I thank Vincent Russo for helpful comments on the writing. This research was supported in part by the National Basic Research Program of China under Awards 2011CBA00300 and 2011CBA00301, and the NSF of the United States under Awards 1017335.

* Electronic address: shiyy@eecs.umich.edu

- [1] E. Simpson, Journal of the Royal Statistical Society, Series B **49**, 238 (1951).
- [2] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
- [3] D. Greenberger, M. Horne, and A. Zeilinger, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, edited by M. Kafatos (Kluwer, Dordrecht, 1989), pp. 69 – 72.
- [4] J. Pearl, *Causality: Models Reasoning and Inference* (Cambridge Univ. Press, 2000).
- [5] J. S. Bell, Physics **1**, 195 (1964).
- [6] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
- [7] B. Cirel'son, Letters in Mathematical Physics **4**, 93 (1980).
- [8] A. J. Leggett and A. Garg, Phys. Rev. Lett. **54**, 857 (1985).
- [9] Č. Brukner, S. Taylor, S. Cheung, and V. Vedral (2004), arXiv:quant-ph/0402127.
- [10] J. Uffink, Phys. Rev. Lett. **88**, 230406 (2002).
- [11] A. Tversky and E. Shafir, Psychological Science **3**, pp. 305 (1992).

- [12] P. Bruza, J. R. Busemeyer, and L. Gabora, Journal of Mathematical Psychology **53**, 303 (2009), special Issue: Quantum Cognition.